1. Grading scales for semiquantitative evaluation for all markers

HLA-DR: immunoreactivity was graded using a five-point scale defined as: 0 (absent immunolabeling), + (few cells), ++ (many microglial cells), +++ (many microglial cells plus focal aggregates / nodules) and ++++ (many microglial cells and macrophages). In the hippocampus and amygdala only scores 0 to +++ were used.

IgG and HIF-1α: immunoreactivity was graded using a four-point scale defined as: 0 (absent immunolabeling), + (few), ++ (intermediate numbers), and +++ (many) immunoreactive cells. A similar scale was used for GFAP in the amygdala.

CD163: immunoreactivity in both, medulla and hippocampus was graded using a five-point scale defined as: 0 (absent labeling), + (few perivascular cells), ++ (numerous perivascular cells), +++ (mainly perivascular cells with few parenchymal cells) and ++++ (many parenchymal cells).

2. Threshold-based analysis of immunostaining by Definiens software

For quantitative analysis of HLA-DR marker in the hippocampus and GFAP marker in the amygdala, sections were initially scanned on a LEICA SCN400F digital slide scanner (LEICA Microsystems, Wetzlar, Germany). Scanned images were processed into a Pyramid TIFF file, stored on a fileserver and viewed and managed on a SlidePath Digital Image Hub (LEICA). Analysis was carried out with Definiens Tissue Studio 3.6 (Definiens AG Munich, Germany). Definiens automated image analysis software identifies 'objects of interest' using a series of pixel-based filters (intensity thresholds and gradients). These filters compare pixels to their neighbours to transform the original image so that the areas of interest can be extracted by simple threshold measures. Prior to the study, a set of 20 sections for each marker (HLA-DR or GFAP) was used as a pilot study to optimise specific parameters for our dataset. First, tissue was segregated from the background of each scanned image by setting

homogeneity to 0.003 units, brightness to 200 units. For HLA-DR marker, intensity threshold was set at 0.12 units to differentiate marked areas from unstained areas. For GFAP marker, overall intensity threshold of marked areas varied significantly (from 0.07 to 0.6 units). Therefore, to differentiate marked areas from unstained areas we divided our dataset into three groups: (i) $\leq$ 0.25 units, low intensity; (ii) 0.25-0.31 units, medium intensity and (iii) 0.32-0.6 units, high intensity. The region of interests were then manually outlined by researchers (ZM for HLA-DR and DO for GFAP) on each image using a stylus pen on a touch-sensitive computer screen. For HLA-DR marker in the hippocampus, eight regions of interest were selected: (i) dentate gyrus (DG), (ii) subgranular zone (SGZ), (iii) CA4, (iv) CA3, (v) CA2, (vi) CA1, (vii) subiculum (Sub) and (viii) parahippocampal gyrus white matter (WM). For GFAP marker in the amygdala, three regions of interest were selected: (i) lateral nuclei (LN), (ii) basal nuclei (BN), (iii) accessory basal nuclei (ABN). After region of interest selection, the intensity thresholds previously set on the pilot groups for HLA-DR and GFAP markers were then applied to all images of every region of interest for the same marker in each case. The data from each region of interest were summarized as a percentage of overall staining (labeling index).